

Weekly Report

2018.0702-2018.0708

1.This Week

Power Grid Deep Learning Paper

1.The data problem discovered last week is solved and now we're successfully generate the data we need. We plan to generate a middle scale training data-set before Chen Zexian leaves so that we can keep training the model when he is out.

2.We compute the effectiveness of each training sample by computing the shanon entropy of it. The computation method is adjusted for many times since there're lots of buses in a sample and the most reasonable result we get is that the entropy of all buses are of the same scale and the buses nearer to the fault center and buses with higher voltage base owns higher entropy.

According to this finding, we use this result to sample informative buses in a sample to get effective training samples to simulate incomplete training data. (We need to do that because in real cases, the incomplete data is received based on the hypothesis that only useful data is provided.) We'll test the model on a small scale of this simulated incomplete dataset (small size and less classes) this week.

SQC Paper

1. Zongzhuang is realizing the CUSUM algorithm with the real case power grid data (data of 36 nodes in the power grid). I asked him to update his work and share with us, but the ddl is passed and I haven't received any thing yet.

Others

- 1.Read machine learning blogs about transfer learning and challenges to common problems.
- 2.revise the waveline paper.
- 3.Make the slice of the power grids projects.

WaveLines Revision:

changes to be made:

- 【done】**
- 1.make clear the definitions of power gris simulation terms like transient simulation stc..(make a form)
 - 2.clearly define all patterns mentioned in this paper and discuss how existing works distinguish these patterns in the related work.
 - 3.revise the evaluation part: add quantitative evaluation of how accurate wavelines can help to find patterns(effentiveness), how long will take(efficiency) and complete understanding (comprehensiveness).

【done】 4.explain the data preprocessing into a whole continuous process.

【done】 5.related work: explain why methods used to visualize pairwise variables mentioned are not used in this paper.(use this to guide the alternative designs discussion.)

6.use less words but add a figure to explain the section waveline design trade-off.

7.explain more carefully why we use a bulb metaphor.

Paper Reading

1.Deep Forest: Towards an Alternative to Deep Neural Networks

This paper proposes a decision tree ensemble approach. This approach is flexible to any scale of data due to its cascade forest structure and its multi-grained scanning process. The experiment presented in this paper suggests this decision tree ensemble approach's performance is competitive to deep neural networks but the training process takes less time. Such lightweight approach can be applied to cases with small-scale training data. But how it would perform with large number of labels/classes cases needs to be tested.

2.Multi-Label Learning with Millions of Labels: Recommending Advertiser Bid Phrases for Web Pages

Machine learning and deep learning classifiers always faces these 2 problems: 1) too many classes; 2) too small data size. Deep forest targets at problem 2 while this paper targets at problem 1. This paper develops a sparse semi-supervised multi-label learning formulation, which is used to infer a belief for each label and the random forest (the model used in this paper is multi-label random forest) classifier is extended to train on these beliefs. The prediction cost is log scale to the number of labels (considered to be efficient).

3.Deep Cocktail Network: Multi-source Unsupervised Domain Adaptation with Category Shift

This paper proposes a transfer learning network that works for labeled source samples coming from diverse sources. It considers the target distribution as a weighted combination of source distributions. The DCP works in two steps: 1) adversarial learning is used to minimize the discrepancy between the target and each of the multiple source domains; 2) the multi-source category classifiers are integrated with the perplexity scores to classify target sample. The experiment result shows this network performs outstanding accuracy with multi-domain sources.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	-	1.Improving the model accuracy. (Solve the data problem and evaluate the value of each training sample.)

Work	Deadline	Progress
SQC Paper	-	About to started.
WaveLine revision	ASAP	Follow the revision plan to revise the paper.